

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-041577

(43)Date of publication of application : 12.02.1999

(51)Int.Cl.

H04N 7/15
G01S 5/18
H04M 3/56
H04N 5/232

(21)Application number : 09-193630

(71)Applicant : FUJITSU LTD

(22)Date of filing : 18.07.1997

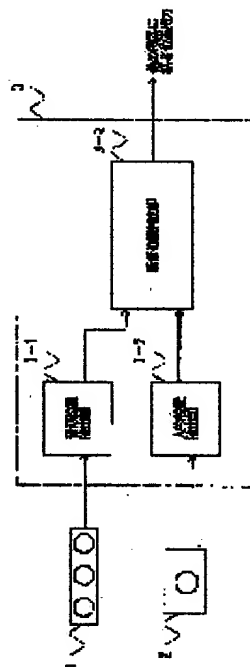
(72)Inventor : KITAGAWA HIRONORI
MATSUO NAOJI
OSADA SHIGEMI

(54) SPEAKER POSITION DETECTOR

(57)Abstract:

PROBLEM TO BE SOLVED: To highly accurately detect a speaker position by judging it while adding two information of the detection information of a sound source position and the detection information of a figure position.

SOLUTION: A sound source position detection part 3-1 of a control part 3 prepares a sound source position map by inputting sound source information inputted from a microphone array 1. Based on the image information inputted by a sensor 2 for image input, a human body position detection part 3-2 prepares a human body position map. The map calculates the probability for the sound source or human body position to exist for the unit of each domain while partitioning the range of a space detectable for the microphone array 1 and sensor 2. A speaker position discrimination part 3-3 calculates the product of probabilities in the correspondent domains of the sound source and human body position maps and discriminates the domain having the largest product as the speaker position. Either one of an ultrasonic sensor, an infrared sensor or a television camera is used for the sensor 2.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-41577

(43) 公開日 平成11年(1999) 2月12日

(51) Int.Cl. ⁸	識別記号	F I	
H 0 4 N 7/15		H 0 4 N 7/15	
G 0 1 S 5/18		G 0 1 S 5/18	
H 0 4 M 3/56		H 0 4 M 3/56	C
H 0 4 N 5/232		H 0 4 N 5/232	C

審査請求 未請求 請求項の数3 O L (全 6 頁)

(21) 出願番号 特願平9-193630

(22) 出願日 平成9年(1997) 7月18日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 北川 博紀

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 松尾 直司

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 長田 茂美

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

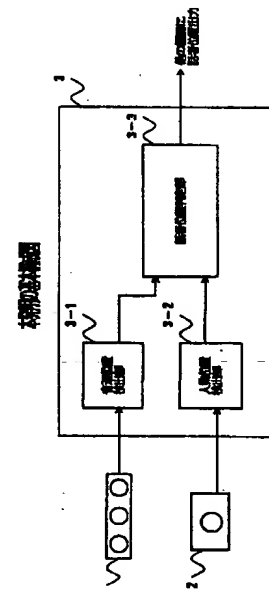
(74) 代理人 弁理士 井桁 貞一

(54) 【発明の名称】 話者位置検出装置

(57) 【要約】

【課題】 テレビ会議システム等でカメラの方向制御に用いられる話者位置の検出を、マイクの位置や人物の位置／動きに関係なく、高い精度で特定出来るようにすることを目的とする。

【解決手段】 マイクロホンアレイの使用により、話者の位置を方向だけでなく、距離も確定して音源位置マップを作ると共に、センサによる画像処理により、人物位置マップを作って、各人の位置を確定する。その2つの情報から、マイクの位置や人物の位置／動きに関係なく、常に話者位置を特定することを可能にする。



【特許請求の範囲】

【請求項1】 テレビカメラ・超音波センサ・赤外線センサ等のいずれか1つのセンサからの入力信号を処理し、人物位置を検出する手段と、マイクロホンアレイからの入力信号を処理し、音源位置を検出する手段と、前記2種類の情報を合わせて処理することにより、話者位置を判定する手段を有し、感知範囲内の任意の位置の話者を検出可能とすることを特徴とする話者位置検出装置。

【請求項2】 請求項1において、センサから見た感知範囲内の水平方向をX、鉛直方向をY、奥行き方向をZとした時、話者位置検出手段として、画像でXY方向の位置を、音源によりZ方向の位置を求め、話者位置を検出することを特徴とする話者位置検出装置。

【請求項3】 請求項1において、センサを回転可能とすることにより、センサの死角に話者がいても、音源位置をもとにセンサを回転させることにより、死角にいる話者の検出も可能としたことを特徴とする話者位置検出装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、テレビ会議等で利用される話者の位置を検出する装置において、話者の位置の検出精度の向上を図った新しい話者位置検出装置に関する。現在、テレビ会議が急速に普及している。テレビ会議においては、カメラを話者に向けたり、カメラの焦点を話者に合わせたりするために、必要な音声のみを拾い出す高精度の話者位置の検出手段が必要とされている。

【0002】

【従来の技術】 従来のテレビ会議装置等で用いられている話者位置検出装置では、各座席に対応して指向性マイクが設置されており、誰かが話すと、最大音量の入力があったマイクの前に座っている人が話者であると確定していた。また、特許出願公開番号「特開平5-244587」に開示された例のように、複数のマイクを水平に等間隔で放射状に配置して並べることにより、話者のXZ平面での方向を特定し、また、カメラから入力した画像を、エッジ検出回路により処理して、輪郭パターンを認識し、予め登録されている人物形状とパターンマッチングを行って人物位置を確定し、更に、動き検出回路により唇の動きを検知して話者位置を特定する例もある。

【0003】

【発明が解決しようとする課題】 しかしながら、上記のような従来例の内、マイクと人を1対1に固定する方式では、話者が移動している場合は対応出来なかった。また別の従来例では、唇の動きを検出するため、話者の正面が見えている必要がある。また、音源位置は、マイクによりXZ平面上での方向しか検出出来なかった。

【0004】 本発明は、話者の位置を特定するために、

マイクと人を1対1に対応させる必要もなく、正面が見えていなくても、或いは、人が移動しながら話をしていても、高い精度で、話者位置を検出することを目的とする。

【0005】

【課題を解決するための手段】 前記目的を達成するために、本発明の話者位置検出装置は、テレビカメラ・超音波センサ・赤外線センサ等のセンサからの入力信号を処理し、人物位置を検出する手段と、マイクロホンアレイからの入力信号を処理し、音源位置を検出する手段と、前記2種類の情報を合わせて処理することにより、話者位置を判定する手段を有し、任意の位置の話者を検出可能とすることを特徴とする。

【0006】 また、前述の装置において、話者位置検出手段として、画像でXY方向のみの位置を、音源によりZ方向のみの位置を求め、話者位置を検出することも出来る。これにより、画像処理が2次元平面上で処理が出来るため、画像処理時間を短縮して、話者位置の検出が可能とすることが出来る。また、前述の装置において、まず、音源位置のXYZ座標を求め、音源位置周辺のセンサ信号のみを処理することにより、音源位置マップの作成と音源位置周辺以外の人物位置検出の画像処理を省くことが可能となり、話者位置検出精度を殆ど落とすことなく、人物位置検出を高速化することが出来る。

【0007】 また、前述の装置において、センサを回転可能とする回転台を設置することにより、センサの死角に話者がいても、音源位置をもとにセンサを回転させ、死角にいる話者の検出も可能とすることが出来る。また、前述の装置において、話者位置のキャリブレーションを行う機能を設けることにより、話者の特定が正しいかどうかを、ディスプレイ等の表示装置上で確認出来るようにし、間違っている場合は、訂正出来るようにすることが出来る。

【0008】

【発明の実施の形態】 本発明の基本構成を、図1を用いて説明する。マイクロホンアレイ1から入力された音源情報を入力にして、制御部3の音源位置検出部3-1が音源位置マップを作成し、話者位置判定部3-3に渡す。画像入力用のセンサ2が入力した画像情報をもとにして、人物位置検出部3-2は人物位置マップを作成し、話者位置判定部3-3に渡す。センサから見た水平方向・鉛直方向・奥行き方向を、それぞれ、X方向・Y方向・Z方向とすると、マップは、マイクロホンアレイとセンサが検出可能な空間の範囲をXYZ方向に一定間隔毎に区切って、各区分単位に音源位置または人物位置の存在する確率を計算したものであり、音源位置マップと人物位置マップは、同じ空間に対応している。話者位置判定部3-3は、音源位置マップと人物位置マップの対応する各区分の確率の積を計算し、その積の最も大きい区分を、話者位置と判定し、話者位置情報を他の機器

に渡す。

【0009】本発明の処理概要を図2のフローチャートを用いて説明する。まず、ステップS1で、音声入力があると、音源位置検出部3-1が、マイクロホンアレイ1の入力信号を分析して、音源位置マップの区分単位に音源位置の確率を計算し、音源マップを完成させる。次に、ステップS2で、人物位置検出部3-2が、センサ2より入力した入力信号を画像処理し、人物位置マップの区分単位に人物位置の確率を計算し、人物位置マップを完成させる。ステップS1とステップS2の処理は、
10 どちらも常時行っており、特にどちらが先というわけではない。2つのマップが完成すると、ステップS3で、話者位置判定部3-3が、音源位置マップと人物位置マップの対応する各区分の積を計算し、その積の最も大きい区分を、話者位置と判定する。話者位置が特定出来ると、ステップS4で、話者位置判定部3-3が、他の機器に話者位置を渡す。

【0010】

【実施例1】図3は、本発明の実施例である。音声が発生すると、マイクロホンアレイ1からの音声情報を、音源位置検出部3-1が、検出範囲内の空間を一定間隔毎に区切り、その区分単位に音源位置の存在する確率を計算し、音源位置マップを作成する。マイクロホンアレイ1からの音声情報により、音源のXYZ座標を求めることが可能である。

【0011】人物位置検出部3-2は、常時、センサ2からの情報をもとに、検出範囲内の空間を一定間隔毎に区切り、区分単位に人物位置の存在確率を求め、人物位置マップを作成する。人物位置の検出は、テンプレートマッチングを用いて検出する。センサ2には、超音波センサ2-2・赤外線センサ2-3・テレビカメラ2-4のいずれか1つを使用し、センサ2は、回転台4の上に載っている。超音波センサ2-2を使用する場合は、超音波発信機2-1と組み合わせて使用する。

【0012】音源位置検出範囲が、センサの感知範囲より広い場合は、話者位置判定部3-3が、回転台制御部3-4に指示して、回転台4を回転させてその回転角度を感知範囲と対応づけ、音源位置検出範囲とセンサの感知範囲を合わせることで、センサの死角を無くす。音源位置マップと人物位置マップは、同一の空間に対応しており、一定間隔毎に区切られている。その各区分は1対1に対応している。

【0013】音源位置マップと人物位置マップの作成が完了すると、話者位置判定部3-3は、マップの各区分単位に2つのマップの確率の積を求め、その積が最大の区分を話者位置と特定する。但し、キャリブレーション機能が、話者位置判定部3-3にある場合は、表示装置6にセンサからの入力画像と話者位置と判定した部分を表示し、人の判断により、話者位置が妥当かどうか判断し、正しい話者位置を入力装置5から入力することが出
50

来る。話者位置判定部3-3は、特定した話者位置の情報を、他の機器に渡す。

【0014】他の機器としては、例えば、テレビ会議システムが考えられ、話者位置情報をもとに、テレビ会議の情報入力カメラの選択・切替え、回転、ズーム等を制御することや、話者の音声のみを強調することや、話者方向にディスプレイを向けること等が可能となる。前述の実施例では、音源位置マップと人物位置マップを完全に作成し、マップの全ての区分について、話者位置の確率を計算しているが、必ずしも全ての区分の確率を計算する必要はない。例えば、マイクロホンアレイを使用した音源位置の検出により、音源位置を求め、センサの人物位置検出を該音源位置から一定の範囲内のマップの区分に絞って検出することにより、話者位置の検出精度を殆ど落とすことなく、高速に処理することが可能である。反対に、センサの人物位置検出を先に行い、検出した人物位置周辺に絞って、音源位置の確率を求め、話者位置を特定することも考えられるが、この場合は、処理時間のかかるセンサによる画像処理を先に行うため、処理の高速化はあまり望めない。

【0015】また、別の方法として、前記のように、マップを3次元の空間を区切って作成するのではなく、処理の高速化のため、人物位置マップは、2次元のXY平面で作成し、音源位置マップは、Z軸の方向を求め、その交点を話者位置と特定することも可能である。

【0016】

【発明の効果】本発明では、マイクロホンアレイによる音源位置の検出情報と、センサの入力信号を処理することによる人物位置の検出情報の2つの情報を合わせて判断することにより、従来より精度の高い話者位置を検出することを可能にすると同時に、従来では出来なかった移動中の話者や画像で唇が検出出来ない場合の話者の特定を可能とした。また、センサに回転台を取り付けることにより、話者がセンサの死角にいる場合も、回転台を回転することにより、話者の位置を確定出来るようにした。

【図面の簡単な説明】

【図1】 本発明の基本構成図

【図2】 処理概要フローチャート

【図3】 本発明の実施例

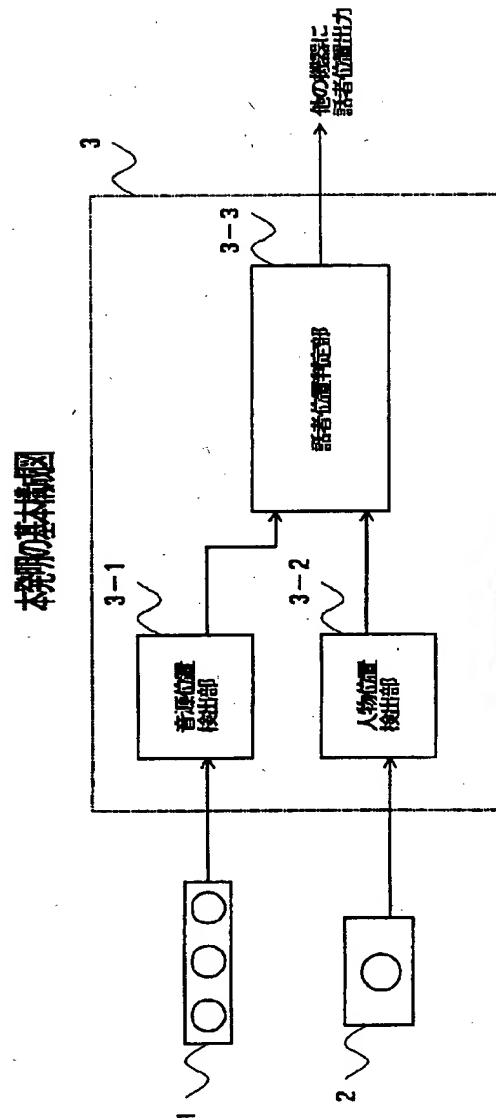
【符号の説明】

- 1 マイクロホンアレイ
- 2 画像入力用のセンサ
- 2-1 超音波発生装置
- 2-2 超音波センサ
- 2-3 赤外線センサ
- 2-4 テレビカメラ
- 3 制御部
- 3-1 音源位置検出部
- 3-2 人物位置検出部

- 3-3 話者位置判定部
 3-4 カメラ制御部
 4 回転台

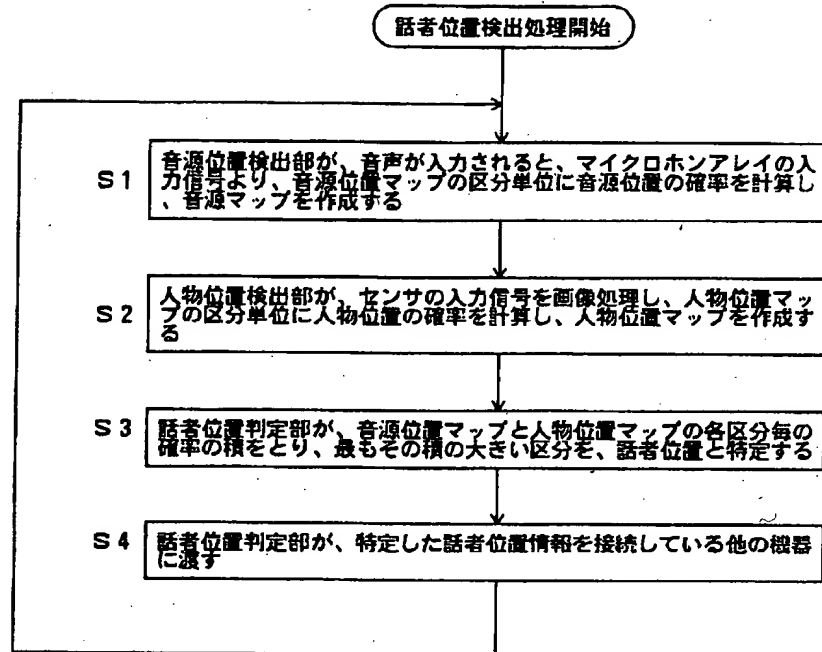
- 5 キーボード・マウス等の入力装置
 6 ディスプレイ等の表示装置

【図1】



【図2】

処理概要フローチャート



【図3】

本発明の実施例

